Selecting and Implementing a Metadata Standard for the Government of Canada

Government On-Line Ad hoc Interdepartmental Metadata Working Group

March 22, 2001

101-04031/

Table of Contents

1	**Why Metadata? What are the benefits?**	**2**
2	**Objectives**	**2**
3	**Why a Metadata Standard Now?**	**3**
4	**Scope**	**4**
5	**Selection Process**	**4**
6	**Selection Criteria**	**5**
7	**The Selection**	**5**
8	**Rules for Use**	**5**
9	**Governance and Leadership**	**6**
10	**Adjunct Activities and Products**	**6**
10.1	Common Vocabulary	6
10.2	Standard Mappings and Crosswalks	7
10.3	Search Engine use of Metadata	7
10.4	Tools	7
10.5	Migration	7
10.6	Training	7
11	**Glossary**	**8**
12	**GAIMWG Members**	**9**
13	**Annex A: Selection Criteria Table**	**10**

1 Why Metadata? What are the benefits?

Metadata is a key tool in describing and managing information assets. Just as we put bar code labels on physical assets such as chairs and computers, or financial codes on our financial transactions, we need a labeling and inventory system to manage information assets. It is particularly important to have an effective identification system for information assets since many are invisible, hidden in web sites or databases, until a user initiates a search to find the assets relevant to a current need. Whether we in the public service or our clients in the Canadian public are the searchers, we need an effective way to use the labels on our information assets to find them when we need them. You can see that the chair you are sitting on is a chair without looking at the bar code label, but an electronic document is invisible until its label or its text is found by a search tool.

The benefits of using a systematic way of assigning and structuring metadata include:

- Reliability: providing information that search engines can use to find relevant documents in large collections such as web sites or document databases where text search alone brings up many irrelevant documents or lists of documents too long for users to look at.

- Identity: providing descriptive information so that users can tell how old a document is, who wrote it, or how to get additional information. Most documents on government web sites now cannot tell the user whether they are 5 days old or 5 years old. Sometimes the user wants one, sometimes the other.

- Inventory: a list of what information the Government holds so that the information can be managed, tracked, updated, analyzed and used efficiently.

- Consistency: an international metadata standard provides the framework and many of the rules for use so that metadata can be applied consistently in large and diverse organizations such as the Government of Canada. This creates an environment in which users can search for and find information without needing to know which department produced it or to which program it relates.

- Interoperability: an international metadata standard provides a way for information resources in electronic form to communicate their existence and their nature to other electronic applications (e.g. via HTML or XML) or search tools and to permit migration of information between applications or search systems.

- Policy compliance: a critical component of meeting the Management of Government Information Holdings (MGIH) policy requirement to know and be able to find the information Government holds.

2 Objectives

The objective of this paper is to raise awareness among senior managers, webmasters and information management practitioners about the need for systematic metadata and the

selection of Dublin Core as a metadata standard with potential for use in diverse government applications. The paper discusses the pressing need for a standard, the process by which a standard has been selected, gaining broader approval, and the work needed to implement the standard.

3 Why a Metadata Standard Now?

The Government On-Line Ad-Hoc Interdepartmental Metadata Working Group (GAIMWG) met on February 27, 2001 to consider the requirement for a metadata standard to serve Government On-Line. Although not unique to GOL, some of the information management issues raised in GOL documentation provided the definition of the problem for which metadata is a key part of the solution:

- Tier 1 Horizontal Enabler: Basic electronic information management guidelines will be developed by Treasury Board Secretariat for departments and agencies to follow when creating, capturing, organizing, displaying, updating, archiving, and disposing of information for their on-line information.

- Tier 2 Expectation: An advanced search capability will be available on the Government of Canada portal and all federal department and agency web sites. Clients will be able to find information and services even if the exact name of the program or service is unknown. Common search principles with similar navigation rules will be implemented across all federal web sites.

- Tier 2 Horizontal Enabler: Treasury Board Secretariat will develop enhanced electronic information management guidelines for departments and agencies to implement. Enhanced information management guidelines will focus on information shared and managed between departments and agencies.

- Tier Three Objective: to promote inter-jurisdictional electronic service delivery. Tier Three builds on Tier Two objectives and applies these same objectives to initiatives with provinces, territories, municipalities, businesses, volunteer organizations, or international partners.

This paper represents the outcome of that discussion on establishing a metadata standard for GOL and the ancillary work needed to accomplish the information sharing, searching and navigation elements of GOL.

The GAIMWG noted that the need for a clearly supported standard and guidelines for its implementation is reaching a critical point. Although the Common Look and Feel (CLF) specifies some mandatory metadata elements for inclusion in web documents, (Common Look and Feel Standard 6.3 < http://www.cio-dpi.gc.ca/clf-upe/6/6_e.asp#6.3 >)the elements do not follow an accepted international standard and there is no clear practical or policy guidance to support the effective and consistent implementation of metadata across departments, or even within departments and programs. As the government's effort to achieve GOL by 2004 picks up speed, it is critical to establish a working standard now to prevent having to retrofit a standard after most GOL and web initiatives have adopted diverse home-grown standards.

Just as globalization affects trade and economic policy, so too, does it affect information management policy. The GAIMWG looked at the practices of other national governments

and the emerging practices in the commercial world. After several years of development in metadata standards, trends have begun to emerge and standards to solidify. The GAIMWG feels that both the need for a standard and the stability-level of international standards are coinciding to make NOW the time to select and implement an international standard. In some departments web site metrics show a growth of 200%-300% per year in the volume of information on government web sites and the rate of use of web sites. With GOL encouraging growth in electronic service and information delivery, this trend is expected to continue. A key component of improving information for electronic resources is the establishment and implementation of a metadata standard for describing and retrieving information.

4 Scope

The metadata standard selected is required urgently for Government On-Line so that information intended to be available to citizens can actually be found by them. Secondary uses would include records management and the management of other primarily internal collections of information. The use of a common metadata standard is important for internal information because it may move into the published domain at later stages in its life cycle or it may be necessary to find unpublished information as part of ATIP processes. Virtually all databases, documents and web sites have metadata associated with them. Establishing a standard will ensure that, across domains and clusters and business functions, the effort that goes into assigning metadata is coordinated, cost effective and client-service focused. The framework in Figure 1. below shows the relationship between the generic international standard and other specialized extensions for specific subject domains or purposes. The GAIMWG supports this framework.

The role of the International Metadata Standard Core: extensible for special subject domains or purposes, but preserves high level consistency and interoperability.

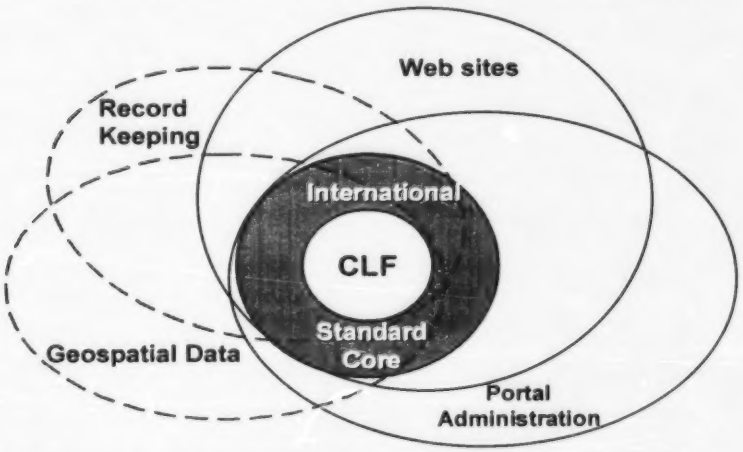

Figure 1. Government of Canada Metadata Framework for Information Resources

5 Selection Process

The GAIMWG membership includes experienced metadata practitioners who have implemented or are implementing metadata in web-based and paper-based information projects. The combined expertise of the group positions it well to select a metadata standard and develop the guidance required to assist departments in implementing the standard. The group was convened to provide direction for GOL metadata efforts. In order for its recommendations to be formalized in a government-wide context or in a broader scope, it will

be necessary to have the GAIMWG standard selection formally reviewed and adopted in senior GOL and IM bodies.

Many departments have indicated that there is an urgent requirement for a standard to be selected and publicized to GOL project teams. Metadata of some description must be attached to GOL information for search and retrieval systems to work well and it makes sense that the efforts expended on metadata be coordinated. If they are not coordinated, there will be a plethora of homemade metadata schemes that will not work together and will require retrofitting to a common standard. For this reason, it is suggested that, in the short term, the GAIMWG metadata standard be adopted as the de facto working standard immediately and that review by senior bodies and the broader community be carried on concurrently with implementation of the standard in GOL projects. The GAIMWG feels that the risk of not having a standard is greater than the risk that their selection would be rejected by senior bodies.

It is recognized that full implementation guidelines and other adjuncts to the implementation of the standard will require considerable effort and discussion in the IM and Web communities. Although we are not as advanced as some other countries in developing such adjuncts, there is sufficient guidance in the international standards documentation to permit the implementation of a basic level of interoperability that will serve the Canadian public well in its information search and retrieval requirements.

6 Selection Criteria

The GAIMWG established criteria for selecting a metadata standard. There are a variety of standards, some of which are well-established in specific subject domains and there are two leading international standards that provide a generic approach to metadata. The two prevailing generic standards are evaluated in Annex A according to the GAIMWG criteria.

7 The Selection

Following the selection process and criteria described above, the GAIMWG has selected the Dublin Core < http://dublincore.org/ >as the most flexible, extensible and current standard, with the firmest footing in technologies of the future.

8 Rules for Use

A metadata standard establishes the semantics or meaning of metadata elements. But no metadata standard can be implemented on a large scale without rules for its use. Just as we now have rules for the use of the Common Look and Feel Elements, we will need to provide departments with rules for the use of the metadata standard. Such rules would include, among other items, the following:

- Scope: when is it appropriate to extend or deviate from the standard? Which elements are mandatory?
- Method: how does one extend or deviate from the standard and still maintain interoperability?
- Content: metadata standards specify the elements, but not what necessarily what words or data go in the elements to describe a document or object. Content rules are needed to

ensure that the content we enter makes sense to those who will search for our information.

- Syntax: does the Government of Canada (GOC) have any special requirements for syntax that are not specified in the existing standard or require a definitive interpretation to specify how the GOC applies the standard?

If rules are not established, the success of this initiative in meeting its goal of facilitating searching, navigation, and information sharing will be jeopardized.

9 Governance and Leadership

In order to be successful as a search tool, a metadata standard for the Government of Canada must have clearly visible support from senior managers and must have a well-functioning governance system. The governance system will need to include:

- Endorsement of a maintenance body or centre of expertise to monitor and update the standard and to issue guidelines and other advice on its application. The maintenance body must include key players in metadata user domains and include members with the necessary training, experience and expertise to assess and make decisions on metadata issues. The maintenance body should also coordinate relations with external communities of interest, such as province, business partners, standards bodies and academic institutions concerned with metadata.
- A senior management body and policy to provide authority for the conduct of the maintenance body.
- An expert advisory group responsible for encouraging the implementation of the metadata standard and for ongoing liaison with and advice to the maintenance body and domains or initiatives that are implementing metadata. This group would participate actively in determining actions on metadata application issues and disseminating information relating to decisions. This group is currently the GAIMWG, which should be recognized with a formal mandate that outlines its full scope and specifies a representative membership.

In addition to a governance system, the successful implementation of a metadata standard requires visible and positive leadership. Such leadership could come from the maintenance body, but would need a senior management champion to lend it authority and credibility and to guide it in its promotion of the standard and the development of tools and services to support implementation of the standard.

Leadership can also come from projects that have been early adopters of the standard and can provide encouragement and guidance to others. SchoolNet is an example of such a project.

10 Adjunct Activities and Products

10.1 Common Vocabulary

As noted above, there must be standards for the content of metadata elements. The metadata standard may say that we should have an element to describe the subject of a document, but it is what each webmaster or content manager enters into the subject field that truly describes the content. We also need to take into consideration the need to be able to retrieve information bilingually and to include both the specialized language of a particular subject area, the language commonly used by bureaucrats, and the language commonly used by citizens. A common vocabulary serves the needs of Canadians searching for information and helps government manage its information assets responsibly. The Depository Services thesaurus meets international standards for multilingual thesauri and covers the wide range of subjects treated by government departments. The DSP thesaurus is endorsed as the Government of Canada Core Subject Thesaurus. Subject specific vocabulary could be

developed or adopted at the department or cluster level in coordination with the core vocabulary.

10.2 Standard Mappings and Crosswalks

It would be convenient to be able to adopt a generic international metadata standard and require that it be used for ALL purposes. However, some specialized subject domains have special needs that no generic standard can fully meet. While trying to keep these deviations to a minimum, we can compensate by mapping standards to each other. Crosswalks are specifications for mapping one metadata standard to another. Standard mappings or crosswalks will need to be made available to permit easy movement between one metadata standard and another. Such mappings and crosswalks have already been created for key international standards and harmonization activities are underway as part of the ISO approval process for metadata standards.

10.3 Search Engine use of Metadata

While commercial search engines that search the Web make only limited use of metadata now, there is a need to encourage more systematic use of metatags. The government also uses search engines to index its own sites. Configuring them to exploit metadata in searches will improve the ability of citizens and public servants to find the specific information they need and to be able to identify important contextual information, such as when a document was written and which department or other author is responsible for the document. To be usable by search engines, metadata must follow a consistent, recognized standard.

10.4 Tools

Tools for tagging documents abound, but there is no standard tool for use in the government of Canada. Existing tools vary in the quality of their interfaces and ease of use. Selection of standard tools would make it easier to assign metadata consistently across government and relieve departments of the burden of trying different tools and developing processes and procedures on their own.

10.5 Migration

Of necessity, departments have added metadata to various kinds of information. This means that there are a number of homemade metadata schemas in use. Where interoperability and ease of access to information is a concern, it will be necessary to develop migration rules, tools and procedures to convert collections to a standard metadata schema.

10.6 Training

Training will be needed for information asset owners, IM practitioners and webmasters to ensure that they understand how metadata is used by clients to find information and by departments to manage their information assets. Training will both inform staff and ensure that there is consistency in the implementation approach taken by departments and clusters.

11 Glossary

Crosswalk	A specification for mapping one metadata standard to another.
Dublin Core	A set of 15 descriptive semantic definitions (metadata elements) that can be shared across disciplines or within any type of organization needing to organize and classify information http://dublincore.org/
GILS	Government or Global Information Locator Service: a set of metadata elements for finding information across collections and disciplines. http://www.gils.net/
Metadata	Information that describes the characteristics of documents or document-like objects to facilitate their retrieval and management. Metadata elements are concerned with the semantics, or meaning, of the metadata tags used.
RDF	RDF specifies the way of encoding metadata tags so a computer can interpret them. http://www.w3.org/TR/REC-rdf-syntax/
Semantics	The meaning of words (or metadata elements).
Syntax	The rules by which words (and metadata elements) can be combined into sentences (expressed in HTML or XML sentences for computers to read).
XML	Extensible Markup Language. In the context of metadata, XML provides the syntax or encoding scheme to express metadata tags in a browser or other computer-readable form. It is the "grammar of choice" for metadata.

12 GAIMWG Members

Name	Department	Phone	e-mail
Chair: Brodie, Nancy	Treasury Board	946-5039	Brodie.Nancy@tbs-sct.gc.ca
Secretary: Miller-Gatenby, Katherine	National Library of Canada	996-5396	katherine.miller@nlc-bnc.ca
Boyko, Ernie	Statistics Canada	951-8218	eboyko@statcan.ca
Brodeur, Jacques	Canadian Heritage	994-5915	jacques_brodeur@pch.gc.ca
Brooks, Diane	Indian and Northern Affairs	994-3257	BrooksD@inac.gc.ca
Clarke, Rachel	Indian and Northern Affairs	953-0161	clarkera@inac.gc.ca
Cowan, Peter	Industry Canada	952-3949	cowan.peter@ic.gc.ca
Curley, Sharron	Public Works and Government Services Canada	956-4847	sharron.curley@pwgsc.gc.ca
Dale, Diana	Canadian Heritage	994-5743	diana_dale@pch.gc.ca
Dulude, Pierre	Canadian Centre for Management Development		pierred@ccmd-ccg.gc.ca
Fontaine, Nora	Department of Foreign Affairs and International Trade	995-6836	nora.fontaine@DFAIT-MAECI.gc.ca
Gnassi, Bruno	Public Works and Government Services Canada	993-3002	bruno.gnassi@pwgsc.gc.ca
Herbert, Lynn	Human Resources and Development Canada	953-2370	lynn.herbert@hrdc-drhc.gc.ca
Hill, Laurie	Statistics Canada (Business Statistics Cluster)		hilllau@statcan.ca
Hjartarson, Fay	National Library of Canada	947-5887	fay.hjartarson@nlc-bnc.ca
Kouri, Bev	Natural Resources Canada	996-3040	bkouri@nrcan.gc.ca
Lilleniit, Roselyn	National Library of Canada	996-7493	roselyn.lilleniit@nlc-bnc.ca
Salloum, Colette	Canada Customs and Revenue Agency		colette.salloum@ccra-adrc.gc.ca
Somerton Cecil:	Fisheries and Oceans		
Steele, James	Human Resources and Development Canada (Benefits Cluster)		james.steele@hrdc-drhc.gc.ca
Tozer, Paula	Environment Canada		paula.tozer@ec.gc.ca
Zeeman, Deane	Finance / Treasury Board		zeeman.deane@fin.gc.ca
Zongora, Catherine	National Archives		czongora@archives.ca

13 Annex A: Selection Criteria Table

Requirement	GILS	Dublin Core
Specialized Domains using		Museums, e.g. Consortium for the Computer Interchange of Museum Information (CIMI); Health, e.g. Catalogue et Index des Sites Médicaux Francophones, WHO;
Adoption by national governments	US for federal government information, although minimally deployed and supported.	Australia[1] (Business Entry Point), Denmark, Finland, UK (for open.gov.uk), Ireland, New Zealand
Implementation track record	Low uptake within US federal government, fewer than 10 states actively adhering. Key barriers: complexity of the standard; expense of implementation	As a "core" standard, DC is not meant to be all things to all people. It is widely perceived as a simpler, cheaper, lighter-weight solution than GILS.
Governance Structure	Weak, low potential for member and user involvement	Strong, enables member and user participation
Maintenance	Weak	Strong international and collaborative maintenance process supported by a body to promulgate changes.
Search engine use	Blue Angel may be integrated with some commercial search engines.	Relationships under development; Blue Angel, Ultraseek, Swish-E, Microsoft's Index Server , Verity Search 97 Information Server
Standards Body Activity	Specified in Z39.50	ANSI
Mappable to other standards (crosswalks)	Yes	Yes
Industry standards support		W3C adoption processes underway (at RFC stage), Open Archives Initiative
Extensible	Yes	Yes
Tools to support metagging available?	Yes	Yes
XML linkage	Weak	Strong
RDF linkage	Weak	Strong
Simplicity for adding tags	Requires professional expertise	Designed for non-professionals
Simplicity from end-user searcher perspective	Use of full element set requires professional expertise; limited element set appropriate for lay user	Designed for use by non-professional
Availability of Training	Low	High
Multilingual support	limited	20 languages

[1] Australia has also developed extensions for records management.

Choix et mise en œuvre d'une norme sur les métadonnées pour le gouvernement du Canada

Groupe de travail interministériel ad hoc du Gouvernement en direct sur les métadonnées

Le 22 mars 2001

Table des matières

1	**Pertinence et avantages de l'utilisation de métadonnées**	**2**
2	**Objectif**	**3**
3	**Pourquoi faut-il dès maintenant adopter une norme?**	**3**
4	**Portée de la normalisation**	**4**
5	**Processus de sélection**	**5**
6	**Critères de sélection**	**6**
7	**Le choix**	**6**
8	**Règles d'utilisation**	**6**
9	**Encadrement et direction**	**6**
10	**Ajustements connexes et produits**	**7**
10.1	Uniformisation du vocabulaire	7
10.2	Normes de mappage et de concordance	7
10.3	Utilisation des métadonnées par les moteurs de recherche	8
10.4	Outils	8
10.5	Migration	8
10.6	Formation	8
11	**Glossaire**	**9**
12	**Membre du Groupe ad hoc**	**10**
13	**Annexe A: Tableau des critères de sélection**	**11**

1 Pertinence et avantages de l'utilisation de métadonnées

Les métadonnées constituent des outils importants pour la description et la gestion des produits d'information. Nous avons besoin d'un système d'inventaire de nos ressources d'information, similaire aux codes à barres apposés sur les produits tangibles, comme les chaises et les ordinateurs, ou aux numéros de confirmation attribués aux transactions financières. Disposer d'un système efficace dans ce domaine s'avère d'autant plus important que les ressources d'information sont le plus souvent invisibles, car elles demeurent enfouies dans une base de données ou un site Web jusqu'à ce qu'un utilisateur lance une requête pour repérer des renseignements dont il a ponctuellement besoin. Tant pour nous que pour notre clientèle, le repérage des renseignements dont nous avons besoin est tributaire d'une utilisation efficace des balises qui décrivent les produits d'information. Pour savoir que la chaise sur laquelle vous êtes assis est une chaise, vous n'êtes pas obligé d'interpréter son code à barres, mais un document électronique reste virtuel tant qu'un outil de recherche ne repère pas son texte ou ses métabalises.

Une méthodologie systématisée d'attribution et de structuration des métadonnées procure différents avantages :

- Fiabilité : La présence d'informations reconnues par les moteurs de recherche peut servir à repérer les documents pertinents dans des corpus d'envergure, comme les sites Web et les bases de données, où la recherche textuelle générerait des listes de résultats trop volumineuses contenant de nombreux documents sans rapport avec la recherche;

- Identification : Il devient possible de fournir à l'utilisateur des renseignements comme la date de création ou l'auteur du document et de lui suggérer des ressources d'information supplémentaires. Sur la plupart des sites Web du gouvernement, l'utilisateur n'est pas en mesure de savoir si les documents repérés datent de cinq jours ou de cinq ans. Or ce peuvent être tout autant les premiers que les seconds qui l'intéressent;

- Inventaire : Disposer d'une liste des informations détenues par le gouvernement facilite la gestion, le repérage, les mises à jour, l'analyse et l'utilisation de celles-ci;

- Cohérence : Une norme internationale fournit un cadre de travail et des règles d'utilisation des métadonnées pouvant s'appliquer de façon uniforme à la grandeur de vastes organisations diversifiées comme le Gouvernement du Canada. Dans l'environnement ainsi créé, l'utilisateur est en mesure de chercher et de repérer les renseignements voulus, même s'il ne sait pas de quel ministère ils proviennent ni de quel programme ils relèvent;

- Interopérabilité : Une norme internationale est un moyen de permettre aux moteurs de recherche et à d'autres applications électroniques (grâce, par exemple, aux langages HTML et XML) de reconnaître la présence et la nature des ressources d'information électroniques et de faciliter la migration de l'information entre ces applications et les systèmes de recherche;

- Respect des politiques : L'utilisation des métadonnées est un élément crucial face aux objectifs de connaissance et de repérage de l'information énoncés dans la politique de Gestion des renseignements détenus par le gouvernement (GRDG).

2 Objectif

Le présent document a pour objectif de sensibiliser les hauts fonctionnaires, les webmestres et les professionnels de la gestion de l'information au besoin d'une structuration systématique des métadonnées et au choix, et éventuellement à l'utilisation dans diverses applications gouvernementales, de la norme du projet « Dublin Core ». Il fait ressortir l'importance de la mise en place d'une norme, le processus qui a mené à la sélection de celle-ci et le travail à réaliser pour sa mise en œuvre.

3 Pourquoi faut-il dès maintenant adopter une norme?

Le Groupe de travail interministériel ad hoc du Gouvernement en direct sur les métadonnées s'est réuni le 27 février 2001 pour définir les caractéristiques d'une norme sur les métadonnées pouvant convenir au projet du Gouvernement en direct. Bien que ne touchant pas exclusivement le GED, certaines questions suscitées dans la documentation du projet ont permis de cerner le problème et de faire ressortir que les métadonnées comptaient parmi les éléments clés de solution :

- Niveau 1 – Déploiement horizontal : Le Secrétariat du Conseil du Trésor développera les grandes lignes directrices en matière de gestion de l'information électronique que les ministères et les agences gouvernementales devront suivre pour la création, l'extraction, l'affichage, la mise à jour, l'archivage et le déclassement de leurs renseignements en ligne.

- Niveau 2 – Préalable souhaitable : Un moteur de recherche évolué devrait être disponible sur le portail Web du Gouvernement du Canada et sur les sites de tous ministères et de toutes les agences gouvernementales. Les clients seraient ainsi en mesure de repérer les renseignements et les services voulus, même s'ils ignorent le nom exact du service ou du programme recherché. Sur tous les sites Web fédéraux, on retrouverait les mêmes principes de recherche et des règles de navigation similaires.

- Niveau 2 – Déploiement horizontal : Le Secrétariat du Conseil du Trésor développera des lignes directrices avancées qui seront mises en vigueur dans les ministères et les agences gouvernementales pour la gestion de l'information électronique. Ces lignes directrices viseront particulièrement les modalités du partage et de la gestion de l'information entre les ministères et les agences.

- Niveau 3 – Objectif : Promouvoir la livraison intergouvernementale des services électroniques. Ce troisième niveau reprend les objectifs du second niveau pour les appliquer à des partenariats développés avec les provinces, les municipalités, les entreprises, les organisations communautaires et nos partenaires internationaux.

Le présent document présente les résultats des discussions sur l'adoption d'une norme sur les métadonnées appropriée aux besoins du GED et sur les travaux accessoires nécessaires pour assurer le partage de l'information et la mise en place du mécanisme de recherche et des éléments de navigation du GED.

Le Groupe de travail ad hoc a constaté que le besoin d'une norme clairement supportée et de lignes directrices pour sa mise en œuvre avait atteint un seuil critique. Même si la Normalisation des sites Internet (NSI) définit quelques métabalises obligatoires pour les pages

Web (Norme 6.3 de la NSI : < http://www.cio-dpi.gc.ca/clf-upe/6/6_f.asp#6.3 >), il s'agit d'éléments de métadonnées qui ne respectent aucune norme internationale reconnue et il n'y a ni encadrement pratique, ni politique formelle pour une utilisation cohérente et efficace des métadonnées entre les ministères, ni non plus entre les programmes au sein des ministères. Comme les travaux du gouvernement s'accélèrent pour terminer le GED d'ici l'an 2004, il devient crucial d'établir dès maintenant une norme fonctionnelle pour éviter les rajustements rétroactifs qui seraient nécessaires si la plupart des projets Internet et du GED élaboraient chacun pour soi leurs propres normes.

Le courant de la mondialisation influe tout autant sur les politiques de gestion de l'information que sur les politiques économiques et commerciales. Le Groupe de travail ad hoc s'est penché sur les pratiques de gouvernements d'autres pays et sur les nouvelles pratiques commerciales ailleurs dans le monde. Au niveau de la normalisation des métadonnées, après plusieurs années de développement, des tendances commencent à émerger et certaines normes se sont stabilisées. Le groupe ad hoc considère que la coïncidence entre notre besoin actuel de normalisation et le niveau de stabilité des normes internationales milite en faveur de l'adoption et de la mise en œuvre d'une norme internationale dès MAINTENANT. Dans certains ministères, les statistiques démontrent que le taux de croissance annuelle de la fréquentation des sites Web et du volume de l'information gouvernementale qu'ils contiennent oscille entre 200 % et 300 %. Avec le GED, qui favorise la croissance de la livraison électronique des services et de l'information, tout laisse à croire que cette tendance va se maintenir. Dans ce contexte, une normalisation des métadonnées pour la description et l'extraction des documents constitue un élément clé pour l'amélioration des ressources électroniques d'information.

4 Portée de la normalisation

Pour que tous les citoyens aient accès à l'information et qu'ils soient en mesure de la consulter, il est devenu urgent d'intégrer au Gouvernement en direct la norme retenue pour les métadonnées. Parallèlement, cette norme pourra aussi s'appliquer à la gestion des dossiers et d'autres corpus de renseignements internes. Au niveau de la documentation interne, l'utilisation de métabalises normalisées prend toute son importance lorsqu'il faut publier des documents dont le cycle de vie se termine ou lorsqu'un processus d'AIPRP exige le repérage d'informations non publiées. Presque tous les sites Web, documents et bases de données sont déjà assortis de métadonnées. L'établissement d'une norme garantira une coordination de leur attribution, tant au niveau des domaines d'application et des groupes de travail qu'à celui des fonctions de gestion, sera efficace et rentable et mettra l'accent sur les besoins du client. La figure 1, ci-dessous, illustre le cadre des relations entre la norme générique internationale et certaines extensions propres à des domaines d'application ou à des fonctions en particulier. Le groupe de travail ad hoc est favorable à l'adoption de ce cadre.

Rôle de la norme générique internationale : Garantir un haut degré de cohérence et d'interopérabilité, tout en offrant la souplesse nécessaire pour accueillir certaines extensions propres à des domaines d'application ou à des fonctions en particulier.

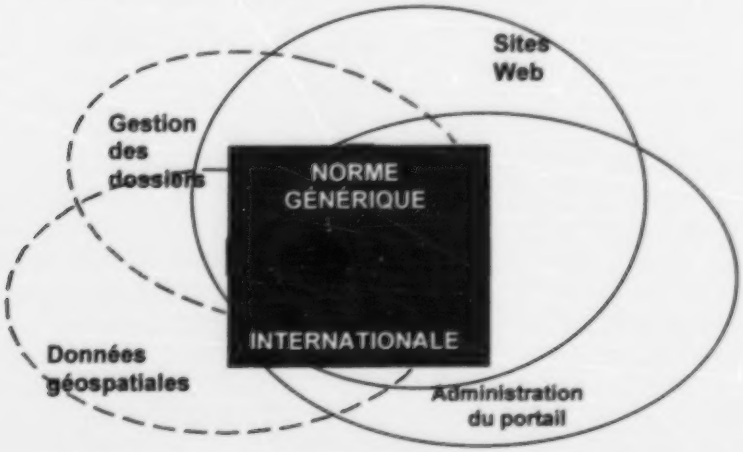

Figure 1. Cadre du Gouvernement du Canada pour l'attribution des métadonnées associées aux ressources d'information.

5 Processus de sélection

Le groupe ad hoc se compose de professionnels du domaine des métadonnées qui ont implanté ou qui implantent actuellement des métadonnées sur des sites Web ou dans le cadre de projets d'information sur support papier. Grâce à cette expertise, le groupe est particulièrement bien désigné pour statuer sur la norme à adopter et pour développer l'encadrement dont les ministères auront besoin pour la mettre en œuvre. Le groupe ad hoc a été mandaté pour conseiller le GED dans le déploiement des métadonnées. Pour officialiser ses recommandations à l'échelle du gouvernement ou dans un contexte plus large, il faudra que la norme retenue par le Groupe ad hoc soit étudiée et formellement adoptée par les cadres supérieurs du GED et les organismes responsables de la GI.

De nombreux ministères ont déjà fait savoir qu'ils considéraient que le choix d'une norme et sa promotion auprès des équipes du projet du GED constituaient une priorité. S'il faut attribuer des métadonnées descriptives aux informations diffusées par le GED pour optimiser les systèmes de recherche et d'extraction des documents, il serait plus sage de coordonner les efforts déployés en ce sens. Sans coordination, on verra apparaître une foule de schémas maison d'attribution des métadonnées qui ne seront pas compatibles entre eux et qui devront par la suite être normalisés de façon rétroactive. C'est pourquoi il est suggéré que, à court terme, la norme retenue par le Groupe ad hoc soit immédiatement adoptée et mise en vigueur et que les organismes décisionnels et la communauté dans son ensemble l'évaluent au cours de son implantation comme norme des projets du GED. Selon le groupe ad hoc, le risque encouru si aucune norme n'est mise en place est plus grave que celui que son choix ne soit pas entériné par les organismes décisionnels.

Il va de soi que la mise en œuvre d'un ensemble complet de lignes directrices pour l'implantation de la norme, de même que des ajustements connexes qui s'imposeront, demandera un effort considérable et de nombreuses discussions dans le milieu de la gestion de l'information et des sites Web. Bien que nous ne soyons pas aussi avancés que dans d'autres pays dans le développement de tels ajustements connexes, on trouve suffisamment d'éléments dans la littérature sur les normes internationales pour s'assurer d'un niveau d'interopérabilité

qui pourra fort bien répondre aux besoins du public canadien en matière de recherche et de d'obtention de renseignements.

6 Critères de sélection

Le Groupe ad hoc a établi une liste de critères de sélection d'une norme sur les métadonnées. Il y a une grande variété de normes. Certaines sont déjà bien reconnues dans des domaines spécifiques d'application, mais deux seulement se démarquent comme chef de file parmi les normes internationales qui proposent une approche globale de la structure des métadonnées. On trouvera dans l'annexe A une évaluation de ces deux normes principales selon les critères établis par le Groupe ad hoc.

7 Le choix

Suite à son processus de sélection, fondé sur les critères indiqués ci-dessus, le Groupe ad hoc a retenu la norme du projet « Dublin Core » < http://dublincore.org/ >, qui lui semble être la plus souple, la plus évolutive et la plus à jour, car c'est celle qui prévoit le mieux l'évolution future des technologies.

8 Règles d'utilisation

Une norme sur les métadonnées établit la sémantique, ou la signification, des métabalises. Il est absolument impossible de mettre en œuvre une telle norme sur une grande échelle sans l'assortir de règles d'utilisation. Tout comme on l'a fait pour la Normalisation des sites Web, il nous faudra fournir aux ministères des règles d'utilisation de la norme sur les métadonnées. Entre autres éléments, ces règles porteront sur :

- La portée : Dans quelles circonstances est-il permis de faire des ajouts ou de déroger à la norme? Quels sont les éléments obligatoires?
- La méthodologie : Quelle est la procédure à suivre pour faire des ajouts ou déroger à de la norme sans compromettre l'interopérabilité?
- Le contenu : La norme sur les métadonnées spécifie les métabalises à utiliser, mais elle ne précise pas nécessairement quel vocabulaire ou quelles données serviront à décrire l'objet ou le document. Des règles de production seront nécessaires pour nous assurer que le contenu des éléments de métadonnées est compréhensible pour les personnes qui interrogent nos bases d'information.
- La syntaxe : Le Gouvernement du Canada (GC) a-t-il des exigences particulières au niveau de la syntaxe qui ne seraient pas prévues par la norme ou désire-t-il une interprétation restrictive spécifiant comment le GED applique la norme?

À défaut d'établir de telles règles, l'atteinte du but de notre entreprise, faciliter la recherche, la navigation et le partage de l'information, serait grandement compromise.

9 Encadrement et direction

Pour que la norme sur les métadonnées retenue par le gouvernement du Canada constitue un outil de recherche qui porte fruit, il lui faut un appui explicite des gestionnaires supérieurs et un système d'encadrement adéquat. Ce dernier doit comprendre :

- La reconnaissance d'un organisme de gestion ou d'un centre d'expertise qui surveillera l'évolution de la norme, la mettra à jour et diffusera des lignes directrices et des avis à propos de son application. Cet organisme devra être constitué de joueurs clés dans le domaine de l'attribution des métadonnées et compter des membres qui disposent de la formation, de l'expérience et de l'expertise nécessaires pour statuer sur les questions relatives aux métadonnées.
- Une politique d'un organisme principal conférant à cet organisme de gestion l'autorité requise pour la réalisation de son mandat.
- Un comité consultatif d'experts chargé de promouvoir la mise en œuvre de la norme et d'assurer la continuité des communications entre l'organisme de gestion et les différents domaines d'application ou projets qui veulent implanter la norme. Ce comité devrait aussi participer activement à la définition des mesures à prendre à propos de toute question relative aux métadonnées et à la diffusion de l'information concernant les décisions qui ont été prises. Le Groupe ad hoc, qui joue actuellement ce rôle, devrait être formellement reconnu comme tel par le biais d'un mandat officiel qui viendrait définir la portée de son action et préciser la représentativité de sa composition.

En plus de ce système d'encadrement, le succès de la mise en œuvre de la norme sur les métadonnées exige une direction visible et positive. Relevant de l'organisme de gestion, cette direction devrait reposer sur l'autorité et la crédibilité d'un gestionnaire principal de haut calibre capable de prendre en charge la promotion de la norme et le développement des outils et des services essentiels à sa mise en œuvre.

La direction pourrait aussi être assumée par des projets qui ont été les premiers à adopter la norme et qui pourraient en encourager et encadrer d'autres. Le Rescol est un exemple de ce type de projet.

10 Ajustements connexes et produits

10.1 Uniformisation du vocabulaire

Comme on l'a souligné plus haut, il faut développer des normes visant le contenu des éléments de métadonnées. La norme générale peut spécifier qu'il faut une métabalise précisant le sujet du document, mais c'est le libellé que chaque webmestre ou gestionnaire de contenu y inscrira qui décrira effectivement le sujet du document en question. Il faut aussi considérer la nécessité d'être en mesure de repérer chaque document dans les deux langues officielles et d'inclure dans les métadonnées tant les expressions plus spécialisées, utilisées par les bureaucrates, que leurs équivalents plus courants dans le langage des citoyens. Le développement d'un vocabulaire commun permet de mieux répondre aux besoins d'information des Canadiens tout en aidant le gouvernement à gérer de façon responsable ses produits d'information. Le thésaurus du Programme des services de dépôt, un ouvrage conforme aux normes internationales visant les thésaurus multilingues, touche une foule de sujets traités par les ministères. Le thésaurus du PDS est reconnu comme Thésaurus des sujets du Gouvernement du Canada. Un vocabulaire spécifique pour certains sujets peut être élaboré et adopté au niveau d'un ministère ou d'un groupe de travail, sous réserve de concordance avec le vocabulaire principal.

10.2 Normes de mappage et de concordance

Il serait fort commode de pouvoir adopter une norme générique internationale sur les métadonnées et d'exiger qu'on la respecte dans **TOUTES** les situations. Mais certains domaines d'application ont des besoins particuliers qu'aucune norme générique ne peut entièrement satisfaire. Tout en tentant de limiter au minimum de telles divergences, on peut compenser les différences en effectuant un mappage des normes. Une table de concordance permet de mapper les unes sur les autres les métadonnées provenant de deux normes

distinctes. Il est donc essentiel de rendre disponibles de telles tables de concordance pour faciliter la conversion des métadonnées entre les normes. De tels outils existent déjà pour les principales normes internationales et les procédures d'harmonisation entre les normes portant sur les métadonnées sont en voie d'obtention d'une certification ISO.

10.3 Utilisation des métadonnées par les moteurs de recherche

Les moteurs de recherche commerciaux ne prenant actuellement que fort peu en compte les métadonnées, il y a un réel besoin de promotion de l'utilisation des métabalises. L'indexation des sites du gouvernement fait aussi appel à ces moteurs de recherche. Une fois que ces moteurs seront configurés pour exploiter des métadonnées, les citoyens, tout comme les fonctionnaires, seront mieux à même de repérer les informations spécifiques dont ils ont besoin et d'accéder à des informations contextuelles, comme la date de création d'un document, son auteur ou le ministère responsable de sa diffusion. Mais pour que les moteurs de recherche puissent efficacement tirer parti des métadonnées, celles-ci doivent satisfaire aux règles d'une norme cohérente et reconnue.

10.4 Outils

Il y a une foule d'outils permettant d'assigner des descripteurs aux documents, mais aucun n'a encore été retenu comme outil standard par le gouvernement du Canada. Le choix d'un outil normalisé faciliterait l'attribution des métabalises à la grandeur du gouvernement et éviterait aux ministères d'avoir à faire l'essai de plusieurs outils et à élaborer des processus et des procédures maison.

10.5 Migration

Par nécessité, certains ministères ont mis en œuvre des schémas de métadonnées visant divers types de document. Cela signifie que l'on utilise déjà plusieurs structures de métadonnées. Si l'on se préoccupe de l'interopérabilité et de la facilité d'accès à l'information, il sera absolument nécessaire d'élaborer des règles, des outils et des procédures pour convertir ces structures afin de les ramener au format défini dans la norme générique.

10.6 Formation

Les propriétaires des produits d'information, de même que les webmestres et les professionnels de la GI, auront besoin d'une formation qui leur permettra de bien comprendre comment les métadonnées sont utilisées par la clientèle en quête de renseignements et par les ministères qui veulent assurer la gestion de l'information. Cette formation permettra à la fois d'informer le personnel et de s'assurer qu'une logique relie l'approche adoptée par les ministères et celle utilisée par leurs équipes de travail.

11 Glossaire

Table de concordance	Ensemble d'instructions permettant de mapper les métadonnées d'une norme sur celles d'une autre norme.
Projet Dublin Core	Le projet de métadonnées « Dublin Core » consiste en 15 éléments de données descriptifs (métabalises) relatifs aux ressources d'information, visant à appuyer le partage de ressources entre diverses disciplines ou au sein de tout type d'organisation qui doit traiter et classifier de l'information. http://dublincore.org/
GILS	(Service de localisation de l'information du gouvernement) Ensemble d'éléments de métadonnées pour le repérage des informations entre diverses disciplines ou collections de documents. http://www.gils.net/
Métadonnées	Données décrivant les caractéristiques d'un document ou d'un objet similaire pour faciliter la recherche, l'extraction et la gestion de l'information. Les éléments de métadonnées représentent la sémantique, ou signification, des métabalises utilisées.
RDF	Resource Description Framework (cadre de définition des ressources) Cadre technique qui spécifie comment encoder les métabalises pour qu'elles soient reconnues et interprétées par l'ordinateur. http://www.w3.org/TR/REC-rdf-syntax/
Sémantique	La signification des mots (ou des éléments de métadonnées).
Syntaxe	Les règles qui permettent d'assembler en phrase des mots (ou des éléments de métadonnées). Pour l'ordinateur, ces phrases sont traduites en langage HTML ou XML.
XML	Extensible Markup Language. Par rapport aux métadonnées, le langage XML fournit la syntaxe ou le schéma d'encodage qui traduit les métabalises en un format reconnu par un navigateur Web ou une autre application informatique. C'est la « grammaire » par excellence pour les métadonnées.

12 Membre du Groupe ad hoc

Nom	Ministère	Téléphone	Courriel
Présidente : Brodie, Nancy	Conseil du Trésor	946-5039	Brodie.Nancy@tbs-sct.gc.ca
Secrétaire : Miller-Gatenby, Katherine	Bibliothèque nationale du Canada	996-5396	katherine.miller@nlc-bnc.ca
Boyko, Ernie	Statistique Canada	951-8218	eboyko@statcan.ca
Brodeur, Jacques	Patrimoine canadien	994-5915	jacques_brodeur@pch.gc.ca
Brooks, Diane	Affaires indiennes et du Nord du Canada	994-3257	BrooksD@inac.gc.ca
Clarke, Rachel	Affaires indiennes et du Nord du Canada	953-0161	clarkera@inac.gc.ca
Cowan, Peter	Industrie Canada	952-3949	cowan.peter@ic.gc.ca
Curley, Sharron	Travaux publics et Services gouvernementaux Canada	956-4847	sharron.curley@pwgsc.gc.ca
Dale, Diana	Patrimoine canadien	994-5743	diana_dale@pch.gc.ca
Dulude, Pierre	Centre canadien de gestion		pierred@ccmd-ccg.gc.ca
Fontaine, Nora	Ministère des Affaires étrangères et du Commerce international	995-6836	nora.fontaine@DFAIT-MAECI.gc.ca
Gnassi, Bruno	Travaux publics et Services gouvernementaux Canada	993-3002	bruno.gnassi@pwgsc.gc.ca
Herbert, Lynn	Développement des ressources humaines Canada	953-2370	lynn.herbert@hrdc-drhc.gc.ca
Hill, Laurie	Statistique Canada (Groupe statistiques d'entreprises)		hilllau@statcan.ca
Hjartarson, Fay	Bibliothèque nationale du Canada	947-5887	fay.hjartarson@nlc-bnc.ca
Kouri, Bev	Ressources naturelles Canada	996-3040	bkouri@nrcan.gc.ca
Lilleniit, Roselyn	Bibliothèque nationale du Canada	996-7493	roselyn.lilleniit@nlc-bnc.ca
Salloum, Colette	Agence des douanes et du revenu du Canada		colette.salloum@ccra-adrc.gc.ca
Somerton Cecil:	Pêches et Océans Canada		
Steele, James	Développement des ressources humaines du Canada (Groupe des prestations)		james.steele@hrdc-drhc.gc.ca
Tozer, Paula	Environnement Canada		paula.tozer@ec.gc.ca
Zeeman, Deane	Finances / Conseil du Trésor		zeeman.deane@fin.gc.ca
Zongora, Catherine	Archives nationales		czongora@archives.ca

13 Annexe A: Tableau des critères de sélection

Critère	GILS	Dublin Core
Domaine d'application spécialisé		Musées (par exemple, le « Consortium for the Computer Interchange of Museum Information (CIMI) »; Santé (par exemple, le Catalogue et Index des Sites Médicaux Francophones, WHO);
Adoption par des gouvernements nationaux	Utilisé aux É.-U pour les systèmes d'information de l'administration fédérale. Le déploiement et le support demeurent toutefois limités.	Australie[1] (« Point d'accès pour l'entreprise »), Danemark Finlande, Royaume-Uni (pour open.gov.uk), Irlande, Nouvelle-Zélande.
Suivi de l'implantation	Démarrage lent aux É.-U.; moins de dix États l'ont adopté. La complexité de la norme et le coût élevé de sa mise en œuvre sont les principaux obstacles.	Comme norme générique, le projet Dublin Core n'est pas conçu comme une panacée. Il est toutefois généralement perçu comme une solution plus légère et supérieure à GILS sur le plan de la simplicité et des coûts.
Structure d'encadrement	Faible. Laisse peu de place à la participation des membres et des utilisateurs.	Forte. Permet la participation des membres et des utilisateurs.
Maintenance	Faible.	Processus coopératif international de maintenance encadré par un organisme qui documente et diffuse les modifications..
Utilisation par des moteurs de recherche	Blue Angel; Peut être intégré à certains moteurs de recherche commerciaux.	Relations d'affaires en développement pour Blue Angel, Ultraseek, Swish-E, Microsoft's Index Server et Verity Search 97 Information Server
Reconnaissance d'organismes de normalisation	Spécifié dans la norme Z39.50	ANSI
Possibilité de mappage vers d'autres normes (concordance)	Oui.	Oui
Support de normes de l'industrie		Processus de certification W3C en cours (à l'étape RFC); Open Archives Initiative.
Évolutivité	Oui.	Oui
Disponibilité d'outils pour la création des métabalises	Oui.	Oui
Compatibilité XML	Faible.	Forte
Compatibilité RDF	Faible.	Forte
Simplicité de l'ajout de métabalises	L'expertise d'un professionnel est requise.	Conçu pour des non-professionnels

[1] L'Australie a aussi développé des extensions pour la gestion des dossiers.

Facilité de la recherche du point de vue de l'utilisateur final	L'utilisation de l'ensemble des possibilités requiert les services d'un professionnel; l'utilisateur final peut mettre en œuvre un éventail restreint de fonctions.	Conçu pour des non-professionnels
Disponibilité de la formation	Faible	Forte
Interface multilingue	Restreinte	20 langues